

NATIONAL INSTITUTE OF ENVIRONMENTAL HEALTH SCIENCES
Division of Extramural Research and Training

NATIONAL ADVISORY ENVIRONMENTAL HEALTH SCIENCES COUNCIL

September 1-2, 2010

Concept Clearance

Statistical, Bioinformatics, and Analytical Methods for Detection of G x E Interactions in
Complex Diseases

Overview:

Much of the burden of health for the U.S. and other developing countries is due to chronic complex diseases such as heart disease, hypertension, diabetes, cancer, asthma, Alzheimer's disease, Parkinson's disease, autism, and many mental health disorders. Multiple genes and environmental factors acting in interconnected biological pathways or networks are thought to contribute to the susceptibility and progression of these complex human diseases. The complexity of the genotype–phenotype relationships in most common complex diseases will require an understanding of genetic and clinical heterogeneity and gene–gene and gene–environment interactions. Genome-wide association studies have emerged as a powerful new tool in recent years for identifying genetic variants to complex diseases. However, most analytical approaches adopted for whole genome scans do not incorporate environmental factors into the analysis and the great majority of genetic variants identified with genome-wide association studies (GWAS) thus far are of little clinical validity since they confer a small increased risk of the disease they associate with, typically with odds ratios of only 1.1-1.5. In addition, some GWAS “hits” appear to be significant only when combined with an environmental exposure. Examples include INSIG2 gene variants with physical inactivity for obesity risk, and TCF7L2 gene variants with dietary fat intake for diabetes risk. The ultimate success of GWAS studies and other gene discovery techniques for contributing to targeted therapies, interventions, or preventive strategies for complex diseases will therefore depend on further development of methods and analytical tools that can incorporate and analyze the role of environment in these diseases. The genetic variants most easily translated for public health or clinical utility will be those that have an obvious environmental modification identified.

Background:

This initiative will encourage the development and validation of algorithms and new statistical and computational approaches to help identify individuals at highest risk for developing a specific disease or dysfunction based on both their exposure patterns and genetic risk profiles. This funding opportunity will also encourage the development of bioinformatics software for G x E analysis of existing populations. These objectives clearly falls within the NIEHS mission and should be a strong component in the NIEHS extramural genomics/environmental health portfolio. A variety of workshops that NIEHS and other NIH Institutes have held in recent years (including multiple GEI sponsored workshops) has stressed the importance of the further development of analytical tools for G x E studies to the field of environmental health sciences.

In addition, a NIH workshop currently planned in September entitled “Next Generation Analytical Tools for Large Scale Genetic Epidemiology Studies of Complex Diseases” will focus on statistical strategies and methods to efficiently identify genetic and environmental factors contributing to complex disease risk. It will also address the resources needed for the design and analysis of large-scale disease association studies and will therefore help to inform this present funding opportunity concept. A defined initiative in this area from NIEHS is long overdue. This funding opportunity also builds on an earlier G x E methods initiative that was a small component of the first phase of GEI.

Objectives: The purpose of this funding opportunity is to develop and test innovative statistical and bioinformatics methods and analytical strategies for identifying gene-environment interactions for complex human diseases. NIH especially encourages the development and application of new tools and computational approaches to further leverage existing human population studies previously utilized for GWAS and other genetic discoveries for G x E interaction studies. Long-term exposures that change over time may indicate certain windows of vulnerability and the temporal and spatial environmental exposure variations need to be incorporated into analytical strategies to appropriately capture exposures that cannot be accurately reflected with just one exposure time point. In addition, the incorporation of prior knowledge about disease pathobiology and gene function is encouraged to computationally model the relationship between combinations of SNPs, other genetic variations, and environmental exposures with disease susceptibility. Some of the possible approaches could include:

- *New statistical methods to identify genetic and environmental interactions that lead to distinct molecular phenotypes (e.g., gene expression).
- *Novel approaches to using biological pathway information in studies of GxE.
- *Approaches allowing efficient G by E analysis across multiple studies.
- *Design and methods for reducing and accounting for measurement error that significantly reduces the power of G by E studies.
- *Novel approaches to efficient sampling designs to increase the range of an environmental exposure to increase power to detect G x E interactions.
- *Methods allowing gene by quantitative exposure interactions.
- *Statistical methods to analyze continuous traits related to a human disease or disorder with multiple genetic and environmental components.
- *Incorporation of systems biology tools to aid in analysis and interpretation of G x E data.
- *Computational methods for data mining and machine learning and bioinformatics methods for incorporating prior biological and toxicological knowledge into data analysis algorithms for identification of G x E interactions.
- *Analytical or bioinformatics tools for data integration (between studies and datasets) and harmonization of environmental data to allow G x E investigations in larger datasets.
- *Computational and bioinformatics approaches to better utilize family studies for G x E discovery.
- *Statistical and analytical methods that allow for incorporation of continuous, accumulating, and/or long-term exposure measurements of environmental exposure(s) combined with multiple genetic components for risk of disease.

Program Management, Implementation, and Requirements:

A program announcement (PAR) with special review is proposed because a specialized study section review is necessary to bring together expertise in the fields of genetics, environmental health, epidemiology, biostatistics, bioinformatics, and computational programming to thoroughly and accurately assess a proposal's ability to move this field forward with novel innovative approaches rather than incremental steps of limited value. A three year R21 will allow exploratory novel approaches to be tested, shared, and made available through open access software. The community will benefit greatly if the investigators of this funding opportunity are required to provide available user-friendly software packages that can be used jointly by researchers with expertise in experimental biology and researchers with expertise in statistics and computer science. Meetings should be held once a year to encourage sharing of statistical methods and approaches for different study designs. Human population datasets must be utilized whenever possible. (The use of simulation data only must be well justified). Non-human models may be needed for some developmental methods (particularly those related to continuous exposure measurements) but this must be justified carefully as well. Program administrators at NHLBI and NCI have expressed interest in participating in this announcement.